# FaceTrust: Assessing the Credibility of Online Personas via Social Networks

Michael Sirivianos
Duke University
msirivia@cs.duke.edu

Xiaowei Yang
Duke University
xwy@cs.duke.edu

Kyungbaek Kim
University of California, Irvine
kyungbak@uci.edu

## 1. Introduction

Nowadays, rich social interactions take place online. Users read, shop, chat, or even play online. Yet the Internet has largely hidden the identity attributes of online users. "On the Internet, nobody knows you are a dog," says the famous Peter Steiner cartoon. When treading through the Internet jungle, what to believe and whom to believe remains a formidable challenge to Internet users. Unscrupulous users may easily become victims of online scams. There have been numerous incidents where scammers defrauded users [3, 13, 15] or even organizations [14] through email or online social networks to obtain sensitive information. Users with vested interest in a company have been caught creating fake positive reviews for the company's products or services [11, 12]. Pedophiles may lie about their ages in online chatrooms, and on the other hand, underage users may also lie about their ages to gain access to age-restricted websites.

This problem stems from the fact that there is currently no lightweight and effective way to assess the *credibility* of assertions/statements made by online personas. In this paper, we refer to credibility as a measure of the likelihood that a user's assertion is correct or true. The current paradigm to address this problem is to use digital credentials issued by trustworthy authorities. Users use them to authenticate at verifying online services. For example, Digital Certification Authorities such as VeriSign [6, 9] can validate a user's identity (name, and domain name) and issue a certificate that assures the user's identity.

These proposals, albeit effective, are heavy-weight and not easily extensible. They often require a tedious registration process that involves manual verification. Obtaining one is not only time consuming but also expensive. As an example, a class one digital certificate from a trustworthy CA costs about $20 [10]. In addition, a digital credential only certifies a limited subset of a user's attributes, such as the Internet domain or name. It does not quantify the general credibility of its holder's assertions, *e.g.*, online reviews or ratings in recommendation sites such as epinions.com.

This paper presents FaceTrust, a system that enables online users or services to assess the credibility of other online personas and their assertions. FaceTrust achieves this goal by mining and enriching information embedded in Online Social Networks (OSNs) such as Facebook [36]. FaceTrust extends an OSN to provide lightweight, extensible, and relaxed digital credentials through *social tagging*. We observe that OSNs already allow users to express a limited form of trust relationships using friend links. FaceTrust extends this ability by allowing users to tag friend links with how credible they consider their friends' assertions (such as their profiles). For instance, a user that wishes to obtain an age certificate from his OSN provider may state that he is above 18 years old in his profile. The OSN provider will request his friends to tag this assertion with a credibility value. The OSN provider analyzes the annotated social graph to obtain the overall credibility of a user's assertion. It can then issue a credential in the form of (assertion, credibility). Websites or other online users may use this OSN-issued relaxed credential to inform their interactions with this user.

FaceTrust's design aims to use social tagging to replace centralized, heavyweight, and restricted verification of user credentials. We face several main challenges in realizing this vision. First, what types of assertions is social tagging able to verify reliably(§2.1)? Second, how can an OSN provider extract the overall credibility information and export it to verifiers without violating a user's privacy (§2.2, 2.4)? Third, how can an OSN provider effectively and efficiently analyze a multimillion-node social graph to extract robust credibility assessments (§2.2, 2.3, 4.2)? Lastly, how can we evaluate the feasibility and performance of such a design (§4)? The main body of this paper describes our initial approaches towards addressing these challenges.

## 2. FaceTrust Design

Figure 1 presents an overview of FaceTrust. The FaceTrust design consists of three main components: a) an OSN provider that maintains the complete social graph and its users' profiles; b) online users that maintain an account with the OSN and attempt to access other online services or communicate with other users by presenting OSN-issued credentials; and c) verifying services or users that regulate access to their resources or characterize user inputs based on their credentials. Next, we describe each component of FaceTrust and how they interact in more detail.

### 2.1 Social Tagging

In FaceTrust, OSN users annotate the social links between them and their friends with additional credibility information. By "social tagging", we refer to the process in which an OSN user assigns a credibility value to a friend or the assertions made by that friend. Tagged values are only known to the OSN and the taggers. Tagging is requested only for users that wish to obtain credentials from the OSN.
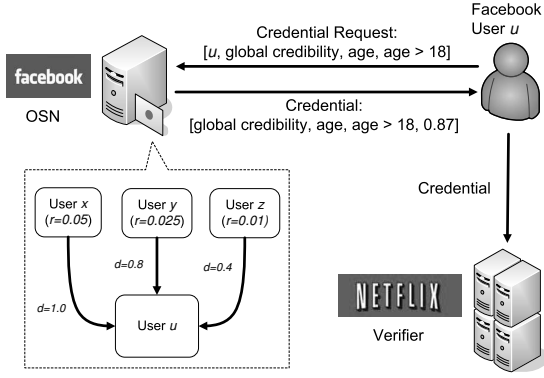
*Figure 1:* **FaceTrust architecture and an age verification example. We use** $d$ **to denote direct credibility and** $r$ **to denote SocialRank.**

FaceTrust relies on social tagging to assess the credibility of online personas. Our assumption is that people usually do not lie on behalf of others. Therefore, the collective information gathered from a user's social acquaintances is likely to correlate positively with the truth. Of course, this assumption does not hold if one user can create multiple fake OSN accounts, an attack known as Sybil attacks [22], For clarity, we assume that each OSN account corresponds to a unique user for the moment, and describe how to mitigate Sybil attacks in § 2.3.

The FaceTrust design categorizes user assertions into different types such as age, address, profession, etc. A user stores his assertions of assorted types in his OSN profile. For instance, for the type age, a user may assert that he is older than 18; for the type profession, he may assert that he is a faculty member at Duke University. For each assertion type $k$ made by a user $j$, $j$'s friend $i$ may tag a direct credibility score as $d_{ij}^k$. Moreover, the user $i$ also tags his friend $j$ with a transitive trust score $t_{ij}$ that indicates how much $i$ trusts $j$ to correctly assess $j$'s friends' assertions of all types.

In our current design, assertion types are mainly identity attributes present in a user's OSN profile, but they can be extended to include other attributes' of users such as fields of expertises. Both direct credibility and transitive trust scores can be discrete values such as "highly trustworthy," "fairly trustworthy," and "untrustworthy" for usability purposes. Presently, we make them continuous values between $[0, 1]$ for ease of analysis.

We note that the FaceTrust design assumes that users are willing to tag their friends. There is abundant evidence that suggests such social tagging may be adopted by users. For example, the "Best-friends" [2] Facebook application enables users to tag and order their friends and has amassed 330$K$ active users. It is our future work to conduct a usability study to validate the adoptability of social tagging.

## 2.2 Assessing Credibility

In the FaceTrust design, an OSN provider plays the role of inexpensive and relaxed credential-issuing authorities. By relaxed, we mean that unlike a conventional certificate au-

thority, the OSN does not guarantee that an assertion is absolutely correct. Instead, each credential is associated with a credibility metric that either resembles "wisdom of crowds" (global credibility § 2.2.1), or the pairwise transitive trust between a verifier and a user (pairwise credibility § 2.2.2). An OSN provider analyzes an annotated social graph to assess the credibility of a user's assertion, and issues a credential that includes the assertion and its credibility score. For instance, in Figure 1, the OSN provider issues a credential that says the global credibility score of the user assertion "age > 18" is 0.87. We describe how an OSN might obtain the global and pairwise credibility assessments in the next.

### 2.2.1 Global Credibility

Let $F_j$ denote the set of friends a user $j$ has. To compute a global credibility score ($g_j^k$) on a user $j$'s assertion of type $k$ ($A_j^k$), FaceTrust computes the weighted average among the direct credibility measures of all $j$'s friends:

$$g_j^k = \sum_{i \in F_j} w_{ij} \cdot d_{ij}^k / \sum_{i \in F_j} w_{ij} \qquad (1)$$

User ratings are weighted by $w_{ij}$ because they are not equally credible, *e.g.*, a teenager's rating on another teenager's age assertion should carry less weight than those from his parents.

How can FaceTrust reliably determine the weights $w_{ij}$? To address this issue, we design a social graph analysis algorithm, SocialRank, to rank the trustworthiness of a user $i$ among all other OSN users based on the transitive trust scores users assign to each other. SocialRank is similar to PageRank [16], which is an iterative algorithm that computes the likelihood that a random walk following the hyperlinks ends up at a page. In SocialRank, we replace PageRank's directed hyperlinks with directed and weighted social links that correspond to the transitive trust assignments from a user $i$ to another user $j$. Let $V$ be the set of nodes in the annotated social graph. Formally, a user $i$'s social rank $r_i$ at each iteration is computed as:

$$r_i = \frac{1 - g}{|V|} + g \sum_{x \in F_i} \frac{r_x \cdot t_{xi}}{\sum_{y \in F_x} t_{xy}} \qquad (2)$$

The parameter $g$ is a damping factor, which we set to 0.9. To initialize the iterations, an OSN provider selects a few ($W$) well-connected users and assigns $\frac{1}{W}$ social rank to each. The computation terminates when the difference between the current social rank vector's norm and the norm in the previous iteration is below a specified threshold.

Intuitively, the higher a user's social rank $r_i$ is, the higher its weight $w_{ij}$ should be. But we can not simply set $w_{ij}$ to $r_i$, because a user with a large number of low-ranked friends may also obtain a high credibility score. To mitigate this issue, FaceTrust assigns a zero weight to a user $i$ if its social rank $r_i$ is in the bottom $b$ percentile of the entire user population. This design assumes that the bottom $b$ percent of users are untrustworthy. For the rest of the users, the weight $w_{ij}$

of a user $i$ in determining another user $j$'s global credibility (Eq. 1) is set to its social rank: $w_{ij} = r_i$.

### 2.2.2 Pairwise Credibility

Global credibility has the drawback that a user's credibility is not directly determined by how much a verifier trusts the friends of the user. In certain cases, *e.g.*, a user $j$ invites a user $v$ to become his friend in an OSN, it may be desirable for the verifier $v$ to use a trusted social path from itself to verify $j$'s assertions. This is a robust technique for $v$ to prevent manipulation from malicious users that collude to boost $j$'s credibility. To this end, FaceTrust's design also includes an algorithm to compute the pairwise credibility $p_{vj}^k$, which assesses the credibility of a user $j$'s assertion $A_j^k$ from the verifier $v$'s point of view.

Pairwise credibility relies on the assumption that trust is largely transitive [25,26]. There have been various proposals to model how trust propagates through a network with a discounting factor, *e.g.*, Guha et al. [26] employ matrix multiplications. Inspired by Credence [35], we use the maximum trust path between the verifier and the user. We choose this approach because the maximum trust path can be computed efficiently using Dijkstra's algorithm.

To compute $p_{vj}^k$, the OSN provider finds the path $P_i$ from $v$ to $i \in F_j$ such that $\Pi_{m \to n \in P_i} \, t_{mn} \cdot d_{ij}^k$ is the largest. The pairwise credibility $p_{vj}^k$ is set to this product. This design assumes that the most useful credibility assessment is the maximum one. Alternatively, to compute $p_{vj}^k$, the OSN provider can find the friend $i$ of a user $j$ that the verifier trusts the most. That is, it finds the path $P_i$ from $v$ to $i \in F_j$ for which $\Pi_{m \to n \in P_i} \, t_{mn}$ is the largest. The pairwise credibility $p_{vj}^k$ is set to $p_{vi} \cdot d_{ij}^k$. This alternative design assumes that the most useful credibility assessment is the one by the most trusted friend of $j$. Our implementation uses the first definition for its simplicity.

### 2.3 Mitigating Sybil Attacks

When malicious users create numerous fake online personas, FaceTrust's credibility assessment can be subverted. For instance, SocialRank is a variation of PageRank and shares its vulnerabilities, but Cheng et al. [19] have shown that PageRank is manipulable using Sybil strategies.

We therefore need a mechanism to combat Sybils when obtaining the pairwise or global credibility. Fortunately, there exist effective algorithms SybilGuard and SybilLimit [38, 39] that can detect Sybil attackers on social graphs. These algorithms take advantage of the feature that most social network users have a one-to-one correspondence between their social network identities and their real-world identities. Malicious users can create many identities or connect to many other malicious users, but they can establish only a limited number of trust relationships with real users. Thus, clusters of Sybil attackers are likely to connect to the rest of the social network with a disproportionately small number of edges.

FaceTrust adapts the SybilLimit algorithm to determine an identity-uniqueness score $q_j$ for each user $j$. The value of $q_j$ is between 0 and 1, indicating the likelihood that an OSN user $j$ corresponds to a unique user in the real life. To be Sybil-resistant, FaceTrust multiplies the identity-uniqueness score $q_j$ to obtain the final global or pairwise credibility score of a user assertion $A_j^k$: $g_j^k \leftarrow g_j^k \cdot q_j$ or $p_j^k \leftarrow p_j^k \cdot q_j$. How to compute $q_j$ is described in a longer version of this work [30].

### 2.4 OSN-Issued Credentials

After an OSN obtains the credibility scores for a user $j$'s assertion $A_j^k$, it can issue a credential for this assertion of $j$. As shown in Figure 1, a credential issued by an OSN will include the assertion type $k$, the assertion $A_j^k$, the type of credibility score, and the credibility score.

A credential must be authenticated by cryptographic primitives such as an OSN's public key signature. In the FaceTrust design, we use the *idemix* [17] anonymous unlinkable credential system because users may desire to preserve their anonymity and untraceability of online activities. The *idemix* system is based on an efficient non-transferable anonymous and unlinkable credential scheme introduced by Camenisch et al. [18]. An *idemix* credential does not reveal any identifying information of a user that possesses the credential, which is ideal for online verifications such as age checking. It also prevents one user from transferring his credentials to other users. More detail on how we integrate *idemix* with FaceTrust can be found in [30].

When a verifier $v$ requests a credential from a user $u$ with a global credibility score, it does not need to reveal any identity information to $u$. However, when a verifier $v$ requests a credential with pairwise credibility from a user $u$, the verifier $v$ must reveal its social network ID to the user $u$ so that the OSN can compute the pairwise credibility $p_{vu}^k$. To prevent revealing its true identity, a verifier should use a pseudonym that by itself does not reveal anything about his real identity, *e.g.*, a Facebook 64-bit semantic-free user ID. In this design, a user $u$ can remain anonymous and unlinkable, but the verifier $v$ can be surveyed if he requests credentials from multiple users. We do not expect that this will become an issue for deployment, because the verifier can choose to request a credential with global credibility to remain unlinkable, or because the multiple users with whom a verifier $v$ interacts may not be aware of each other.

## 3. Sample Applications of FaceTrust

Next, we briefly describe how FaceTrust's global and pairwise credibility can be used in various scenarios. Unlike Certificate Authorities [6] or access control systems such as Kerberos [32] and Shibboleth [23], OSN-issued credentials can not be completely trusted. However, we argue that in many real world situations, they provide valuable and otherwise unavailable information for access control and identity verification.

## 3.1 Age Verification

A concrete example of FaceTrust credentials is age verification in online settings. Figure 1 shows an example. User $u$ attempts to access an age-restricted movie at the Netflix website. At the same time, $u$ is concerned with his anonymity and does not wish to reveal neither his real identity nor a linkable pseudonym to Netflix.

Since Netflix as an entity may not be a user of an OSN, it will demand an age crendential with a global credibility score. To obtain this credential, user $u$ issues a credential request for global credibility of assertion type "age". This request typically happens before $u$ attempts to access an online service. User $u$'s friends on the OSN who have seen this assertion can judge whether it is correct, and tag this assertion with a credibility score.

In Figure 1, there are three friends that have tagged $u$'s age credibility. These friends, users $x$, $y$, $z$, have social ranks equal to 0.05, 0.025 and 0.01 respectively. In addition, users $x$, $y$, and $z$ assign direct credibility scores to $u$'s age assertion ("I am 21"): 1.0, 0.8 and 0.4, respectively. The global credibility of $u$ is derived using Equation 1. At this point, the OSN provider considers that there is 0.87 probability that $u$'s age assertion is correct, assuming that $u$'s identity uniqueness has been determined to be 1.

The OSN compares the assertion for which the credential is requested ($age > 18$) with the assertion $u$'s friends have tagged. It determines that the requested assertion does not contradict the assessed one ($21 > 18$). Consequently, the OSN issues the credential depicted in Figure 1 and $u$ presents it to Netflix. Netflix uses this credential to decide whether to allow $u$ to download the movie.

## 3.2 Fraud Detection

FaceTrust can help a user to detect malicious intentions when he is contacted by an unknown user. Scammers commonly respond to online postings alleging to be prospective participants in legitimate transactions but in reality aiming to commit "advance-fee" fraud [7]. Misbehaving users may create fake OSN personas to defraud other users [13, 15]. Such attacks could be averted if scammers were unable to lie about their location, affiliation, age,

In this example, suppose a user $v$ receives a message from $u$ in response to his Craigslist posting offering a room to rent. Both users are members of an OSN that supports FaceTrust. User $u$ is actually a scammer residing in a suspicious country. On the other hand, $u$ claims in his message that he currently resides in US.

With FaceTrust, user $v$ can require $u$ to present a credential on the pairwise credibility regarding $u$'s location: $p_{vu}^{loc}$. If $v$ is not worried about other users tracking his activities (because other users may not know each other), he can release to $u$ his semantic-free OSN ID. User $u$ then requests from the OSN a pairwise credibility credential with respect to $v$. For this example, we assume that the maximum trust path is $v \rightarrow x \rightarrow y \rightarrow u$, and the transitive trust scores $t_{vx}$ and $t_{xy}$ are 1.0

and 0.5, respectively. Also, $y$'s direct credibility rating on $u$'s location $d_{yu}^{loc}$ is 0.3. According to § 2.2.2, the pairwise credibility $p_{vu}^{loc}$ is 0.15. Subsequently, the OSN issues the following credential for $u$:

[*pairwise credibility*, *location*, *USA*, 0.15]

User $u$ can then present the credential to $v$. Likely, the user $v$ will be alarmed by $u$'s low credibility and refrain from any further dealings with $u$.

# 4. Preliminary Evaluation

To gain a better understanding on how our initial design works, we would like to evaluate all of the following aspects of FaceTrust:

- Effectiveness: How well do credibility scores correlate with the truth?
- Computational feasibility: A social network may consist of several hundreds of millions of users. Will an OSN provider have sufficient computational resources to mine the social graph and derive credibility measures?
- Robustness: How robust is the design in withstanding various malicious attacks or incorrect user tagging?
- Usability: How often and how accurate will a user tag his friends to help them obtain credentials?

It will take a full system implementation and experimentation on a real-world OSN to answer these questions. The question regarding effectiveness is particular difficult, because trust is inherently subjective, and it might not even be feasible to obtain the ground truth. In this section, we describe our preliminary approaches to evaluate the effectiveness and feasibility of the design. We defer the robustness and usability study to future experimentations on an OSN such as Facebook. We have also evaluated how well the FaceTrust design withstands Sybil attacks. Results are omitted here (but available in [30]). They agree well with previous studies that show the effectiveness of SybilLimit [38].

## 4.1 Effectiveness

FaceTrust's design uses the SocialRank algorithm to derive global credibility, and the maximum trusted path to derive pairwise credibility. As the SocialRank algorithm is similar to PageRank, which has been shown to be effective in ranking pages regarding their relevance to users, we assume that it is also effective in ranking users they are trustworthy. Real-world experimentation is required to validate this hypothesis, *i.e.*, to verify whether the highly ranked users are indeed more likely to make a correct rating than lowly ranked ones. We defer this study to future work.

As a first step, we evaluate the effectiveness of pairwise credibility (§ 2.2.2) in verifying identity assertions made by online personas, using a well-known and deployed social graph that embed identity assertions: the PGP Web of Trust (WoT) [4]. In WoT, each user $x$ corresponds to a unique public key. Each edge between $x$ and another user $y$ corresponds

| Orig. Trust Level | 1.0 | 0.5 | 0.0 |
|---|---|---|---|
| Exact (%) | 92.6 | 22.9 | 68.1 |
| Fair (%) | 1.7 | 70.2 | 3.7 |
| Wrong (%) | 5.7 | 6.9 | 28.2 |

*Table 1:* Prediction performance of transitive trust.

to a signature by $x$ on $y$'s public key. The signature includes an annotation indicating the level of trust $x$ places on $y$, both in terms of $y$'s identity and in terms of $y$'s ability to evaluate the identity of others.

Recall that pairwise credibility is derived from the maximum transitive trust. Thus, we use the WoT graph to study how effective the maximum transitive trust is in predicting the true trust level between two users. To do so, we randomly remove an edge between a node $x$ and $y$ ($x \rightarrow y$) in the WoT graph. We then compute the maximum transitive trust $p_{xy}$ as described in § 2.2.2 between $x$ and $y$, and compare $p_{xy}$ with the original trust level $t_{xy}$. If they match, it indicates that transitive trust is effective in predicting the truth.

The WoT data we use are PGP key certificates stored at the Swiss keyserver [8]. The data set represents a snapshot of the WoT between July 19 2008 and July 21 2008, and includes 39625 keys and 398660 signatures. The trust levels specified in a certificate have four discrete values: unknown, untrusted, marginal, and full. We map the discrete trust levels to credibility values in $[0,1]$ to facilitate transitive trust computation. The level "unknown" and "untrusted" are both mapped to 0; the level "marginal" is mapped to 0.5; and the level "full" is mapped to 1.0. The percentages of full, marginal, and unknown/untrusted edges in the WoT graph are 29.47%, 6.8%, and 63.73%, respectively.

Table 1 depicts how well transitive trust predicts the original trust level for each removed edge $x \rightarrow y$. The results are averaged over 1000 randomly removed edges for each trust level. Note that if a node $y$ has only one incoming edge $x \rightarrow y$, we do not remove that edge because $y$ will be disconnected from the WoT graph and we can not use transitive trust to predict the trust level on $x \rightarrow y$. In the table, "Exact" means that the maximum transitive trust $p_{xy}$ between nodes $x$ and $y$ is exactly the same as the original trust value $t_{xy}$ on edge $x \rightarrow y$. "Fair" means that the obtained pairwise trust value has fairly accurately predicted the original trust level. For the full trust level 1.0, a prediction above or equal 0.5 is considered a fair match; for the marginal trust level 0.5, a prediction above 0.5 is considered fair; for the untrusted/unknown trust level 0.0, any value below or equal to 0.5 is considered fair. Predictions in all other ranges are considered "Wrong."

We observe that maximum transitive trust predicts the original trust level satisfactorily. For the original trust level 1.0 and 0.0, the transitive trust prediction has 92.6% and 68.1% exact matches respectively. For the original trust level 0.5, the transitive trust prediction yields 70.2% marginal matches. These results suggest that pairwise credibility computed using transitive trust is likely to predict the level of credibility that a user $x$ itself would directly assign to a user $y$, if $x$ and $y$ were acquainted.

## 4.2 MapReduce Evaluation On Feasibility

Next, we evaluate whether OSN providers that typically employ clusters of thousands of machines can efficiently analyze large social graphs and provide FaceTrust's credential service. To this end, we implement the algorithms to compute pairwise credibility and identity uniqueness for Sybil-resistance using MapReduce [21], and test the efficiency of the implementation using a 500K user crawled Facebook social graph sample, which we obtain from a previous study [24]. We also sample the crawled Facebook graph using the "forest fire" [27] sampling technique to create social graphs of different sizes to study the implementation's scaling factor.

MapReduce is a programming framework for performing distributed processing of large data sets using a large number of machines (nodes). We use the EC2/S3 [1] cluster and the Hadoop [5] Java-based MapReduce framework. Our cluster uses at most 20 machines at any time. The MapReduce implementation is described in [30]. We have not evaluated the SocialRank algorithm because it is based on PageRank whose feasibility has been thoroughly proven in practice.

We evaluate the time needed to compute the pairwise credibility between all pairs of users. The complexity of this computation on a credibility-annotated social graph $S(V, E)$ is $O(|V|^2 log|V|)$. The computational cost increases proportionally to the square of the network size. Therefore it is CPU bound, rather than memory bound.

In Figure 2(a), we observe that the computational cost grows dramatically as the network size increases. It takes close to 50 hours to finish the pairwise credibility computation on a 500K user graph. Although the completion time is quite large in our small 20-node cluster, we think that it is likely to be much smaller for a real-world OSN provider, and thus the social graph analysis algorithms are feasible. This is because OSN providers will typically have hundreds of times higher computational capacity than our profiling testbed, and the computation time decreases almost linearly with the increased number of machines (Figure 2(b)). With a 2000-node cluster, the computation time for a 500K graph is likely to be $\sim$ 30 minutes, shortened by 100-fold. In addition, an OSN only needs to perform the social graph analysis periodically (such as daily) as a background task because the social graph structure is likely to remain largely unchanged in a short period of time. Furthermore, OSN providers can partition multimillion-user social networks into subgraphs of more manageable size (*e.g.*1 million), by exploiting OSN groups or geographic networks. Finally, the all-pair maximum trust algorithm could be optimized for sparse graphs to further reduce the computation time, a step we are currently pursuing.

We have also evaluated the efficiency of the identity uniqueness computation (§ 2.3) for Sybil attack mitigation. This computation is more efficient than pairwise credibility because the complexity of the our MapReduce SybilLimit-based
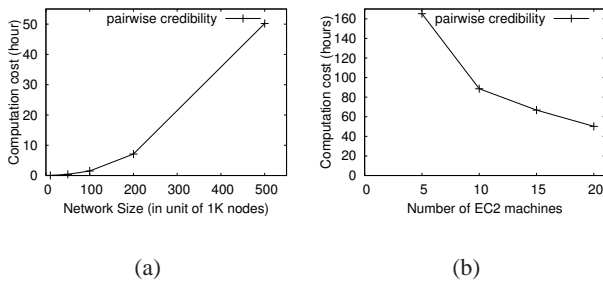
*Figure 2:* **(a) Computation time to derive pairwise credibility for all pairs of users as a function of the number of users in social graphs of various sizes; (b) Computation time to derive pairwise credibility for all pairs as a function of the number of machines in the EC2/S3 cluster.**

implementation is $\Theta(|V| * \sqrt{|E|})$. Results are included in [30].

## 5. Related Work

The goal of FaceTrust is mostly related to PGP Web of Trust [40]. Like PGP, FaceTrust aims to circumvent the expensive and often monopolized Certificate Authorities such as VeriSign to provide lightweight credentials. Unlike PGP, FaceTrust uses the intuitive OSN interface, and employs social tagging rather than key signing to obtain trust metrics. Furthermore, FaceTrust is easily extensible, and is not limited to certifying only public keys. Users can tag each other regarding multiple types of identity attributes, and this set can be extended by adding fields into a user's profile. FaceTrust also uses *idemix*'s credential design to provide anonymity and unlinkability, features that PGP does not provide.

FaceTrust adapts several link analysis algorithms from previous work such as PageRank [16], transitive trust analysis algorithms [35], and Sybil-resistance algorithms [38]. However, our contributions are not the graph algorithms per se. Instead, they lie in the novel idea of using OSNs to provide lightweight, extensible, and relaxed credentials, and the overall design and preliminary evaluation of FaceTrust. A variety of systems have employed trust in social networks to improve system security [20, 28, 29, 31, 33, 34, 37]. To the best of our knowledge, this is the first work that proposes to use OSNs to provide credentials for online personas.

## 6. Conclusion

Despite the large volume of social interactions taking place on the Internet, it is still hard to assess the credibility of statements made by online users. This paper presents FaceTrust, a system that leverages online social networks to provide lightweight, flexible, and relaxed credentials that enable users to assess the credibility of others and their assertions. In the FaceTrust design, OSN users explicitly tag transitive trust and credibility scores to their friends and their identity assertions available in their social network profiles. An OSN provider analyzes the annotated social graph to assess the credibility of a user's identity assertions, and issue credentials annotated by credibility scores. Our preliminary evalu-

ation suggests that FaceTrust is feasible and could be effective in obtaining credible and otherwise unavailable identity information for online personas.

## 7. References

[1] Amazon elastic compute cloud (ec2). http://aws.amazon.com/ec2/.
[2] Best friends facebook application. http://www.facebook.com/applications/Best_Friends/6421213170.
[3] Craigslist scams. http://www.craigslist.org/about/scams.
[4] GNU Privacy Handbook. http://www.gnupg.org/gph/en/manual.html#AEN385.
[5] Hadoop. http://wiki.apache.org/hadoop/HadoopMapReduce.
[6] Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile.
[7] Nigerian Advance Fee Fraud. www.state.gov/www/regions/africa/naffpub.pdf.
[8] Swiss keyserver Web of Trust.
[9] VeriSign, Inc. http://www.verisign.com.
[10] Verisign: Personal Digital Certificate Enrollment. https://personalid.verisign.com.au.
[11] Belkin's Amazon Rep Paying For Fake Online Reviews. http://hardware.slashdot.org/article.pl?sid=09%2F01%2F17%2F166226&from=rss, 2009.
[12] Carbonite Stacks the Deck With 5-Star Reviews. http://tech.slashdot.org/article.pl?sid=09/01/27/2349217, 2009.
[13] Fears of Impostors Increase on Facebook. http://www.cnn.com/2009/TECH/02/05/facebook.impostors/index.html?eref=rss_topstories, 2009.
[14] Nigerian Accused in Scheme to Swindle Citibank. www.nytimes.com/2009/02/21/nyregion/21scam.html?_r=3, 2009.
[15] Teen Accused of Sex assaults in Facebook Scam. http://www.msnbc.msn.com/id/29032437/, 2009.
[16] S. Brin and L. Page. The Anatomy of a Large-scale Hypertextual Web Search Engine. In *Computer Networks and ISDN Systems*, 1998.
[17] J. Camenisch and E. V. Herreweghen. Design and Implementation of the idemix Anonymous Credential System. In *ACM CCS*, 2002.
[18] J. Camenisch and A. Lysyanskay. An Efficient System for Non-transferable Anonymous Credentials with Optional Anonymity Revocation. In *EUROCRYPT*, 2001.
[19] A. Cheng and E. Friedman. Manipulability of PageRank under Sybil Strategies. In *NetEcon*, 2006.
[20] G. Danezis and P. Mittal. SybilInfer: Detecting Sybil Nodes using Social Networks. In *NDSS*, 2009.
[21] J. Dean and S. Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. In *OSDI*, 2004.
[22] J. R. Douceur. The Sybil Attack. In *IPTPS*, March 2002.
[23] M. Erdos and S. Cantor. Shibboleth Architecture DRAFT v05. *Internet2/MACE, May*, 2, 2002.
[24] M. Gjoka, M. Sirivianos, A. Markopoulou, and X. Yang. Poking Facebook: Characterization of OSN Applications. In *WOSN*, 2008.
[25] Gray, Seigneur, J.-M., Y. Chen, and Jensen. Trust propagation in small worlds. In *LNCS*, pages 239–254. Springer, 2003.
[26] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of Trust and Distrust. In *WWW*, 2004.
[27] J. Leskovec and C. Faloutsos. Sampling from Large Graphs. In *SIGKDD*, 2006.
[28] A. Mislove, A. Post, P. Druschel, and K. P. Gummadi. Ostra: Leveraging Social Networks to Thwart Unwanted Traffic. In *NSDI*, 2008.
[29] A. Ramachandran and N. Feamster. Authenticated Out-of-Band Communication Over Social Links. In *WOSN*, 2008.
[30] M. Sirivianos, X. Yang, and K. Kim. FaceTrust: Assessing the Credibility of Online Personas via Social Networks.
[31] Y. Sovran, A. Libonati, and J. L. Pass it on: Social Networks Stymie Censors. In *IPTPS*, 2008.
[32] J. Steiner, C. Neuman, and J. Schiller. Kerberos: An Authentication Service for Open Network Systems. In *USENIX Winter Conference*. Dallas), 1988.
[33] A. Tootoonchian, K. K. Gollu, S. Saroiu, Y. Ganjali, and A. Wolman. Lockr: Social Access Control for Web 2.0. In *WOSN*, 2008.
[34] D. N. Tran, B. Min, J. Li, and L. Subramanian. Sybil-Resilient Online Content Rating. In *NSDI*, 2009.
[35] K. Walsh and E. G. Sirer. Experience with an Object Reputation System for Peer-to-Peer Filesharing. In *NSDI*, 2006.
[36] Facebook. http://www.facebook.com, 2008.
[37] S. Yardi, N. Feamster, and A. Bruckman. Photo-Based Authentication Using Social Networks. In *WOSN*, 2008.
[38] H. Yu, P. Gibbons, M. Kaminsky, and F. Xiao. A near-optimal social network defense against sybil attacks. In *IEEE S&P*, 2008.
[39] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. SybilGuard: Defending Against Sybil Attacks via Social Networks. In *ACM SIGCOMM*, 2006.
[40] P. Zimmerman. PGP Users Guide. December 1992.